

# 人格测验中条目呈现方式与记分方式的效应初探

罗杰<sup>1</sup>, 戴晓阳<sup>2</sup>

(1. 贵州师范大学教育科学学院, 贵州省普通高校基础心理与认知神经科学特色重点实验室, 贵阳 550001; 2. 深圳大学心理系, 深圳 518060)

**【摘要】 目的:**初步探讨人格测验中条目呈现方式和记分方式的效应。**方法:**采用2×2研究设计,运用《大五人格形容词评定词表》对734名在校大学生进行测试。**结果:**①无论6点还是7点记分,双极词的因子结构均优于单极词的结果;②6点记分双极词的因子结构较优于7点记分,而7点记分双极词的 $\alpha$ 系数略高于6点记分的结果,在重测信度上两种记分方式的结果比较接近。**结论:**不同的条目呈现方式和记分方式对人格测验的心理测量学性能存在影响,测验编制者在建构或开发测验工具时应尽量选择合适的测验条目与记分方式,以达到事半功倍的效果。

**【关键词】** 人格测验; 单一-双极词; 记分方式; 心理测量学属性

中图分类号: R395.1

DOI: 10.16128/j.cnki.1005-3611.2015.02.011

## The Effects of the Item-presenting and Item-scoring in Personality Measurement

LUO Jie<sup>1</sup>, DAI Xiao-yang<sup>2</sup>

<sup>1</sup>School of Educational Science, Guizhou Normal University, Guizhou General Colleges Key Laboratory of Fundamental Psychology and Cognitive Neuroscience, Guiyang 550001, China;

<sup>2</sup>Department of Psychology, Shenzhen University, Shenzhen 518060, China

**【Abstract】 Objective:** To explore the effects of the item-presenting and item-scoring in personality measurement. **Methods:** This study applied a 2×2 two factor experiment design, 734 university students were surveyed with the big five-factor personality trait-descriptive adjectives scale. **Results:** ① Whatever it was 6-point or 7-point, the bipolar variables provided more robust and univocal factor structure than unipolar counterpart. ② The 6-point bipolar scales provided more robust and univocal factor structure than the 7-point bipolar counterpart; additionally, with this format the coefficient  $\alpha$  estimate was a little lower than the 7-point bipolar counterpart, but the test-retest reliability coefficients of the two formats were similar. **Conclusion:** The different item-presenting and item-scoring influence the psychometric properties of personality measurement, and the appropriate item-presenting and item-scoring should be taken into account while developing the personality measurement, the viewpoints of researcher.

**【Key words】** Personality measurement; Unipolar-bipolar variable; Item-scoring; Psychometric properties

有关人格测验中条目呈现方式的争议主要表现在究竟哪种形式的条目更为简便,更容易让被试理解,同时又能够获得较好的测量结果。以大五人格测验为例,当前国际上主要存在三种形式的自陈式大五人格测验:即句子式、形容词式和短语式<sup>[1,2]</sup>。其中,句子式的条目呈现方式有正向表述和反向表述两种,有研究<sup>[3-9]</sup>发现不同的条目表述方式对测验的信效度存在影响。而形容词式的条目编排分为两种:单极词(如外向的)和双极词(如外向的一内向的)。相对于单极词来说,双极词能更清晰地表述量表所要测量的维度和内容<sup>[10]</sup>。使用双极词的优点在于能使得被试更容易理解问题,其不足之处在于运用双极词时就使得每对双极词的两端都被强制看成是相反的,尽管对所有的被试来说它们并不完全是一对反义词<sup>[11]</sup>。Goldberg通过比较这两种条目呈现方式所编制测验的因子结构,发现单极词的因子稳

定性优于双极词<sup>[12]</sup>。故目前在国外既有单极词形式的人格测验,如100-TDA(100 unipolar Trait Descriptive Adjectives)<sup>[12]</sup>;也有双极词形式的人格测验,如50-BRS(50 Bipolar Adjective Rating Scales)<sup>[12]</sup>,23BB5(23 Bipolar Big Five questionnaire)<sup>[11]</sup>和BFQ(Big Five Questionnaire)<sup>[13]</sup>。

社会与行为科学领域中,测验工具的记分方式主要包括0、1记分和李克特多级记分。其中前者主要有EPQ、MMPI等较早开发的人格量表,而目前大多数研究者更倾向于采用李克特多级记分。其中句子式人格测验多采用5点或6点记分,如NEO-PI-R原版及修订本<sup>[14,15]</sup>、HEXACO-PI(Honesty-Humility Emotionality Extraversion Agreeableness Conscientiousness Openness to Experience Personality Inventory)<sup>[16]</sup>、QZPS(Chinese Personality Assessment Scale)<sup>[17]</sup>和QZPS-Q(Chinese Adolescent Personality Scale)<sup>[18]</sup>等均采用5点记分;CBF-PI(Chinese Big Five Personality

Inventory)<sup>[19]</sup>则采用6点记分。而形容词式人格测验则大多采取7点或9点记分,前者如23BB5<sup>[11]</sup>、BFQ<sup>[13]</sup>等;后者50-BRS<sup>[12]</sup>、100-TDA<sup>[12]</sup>和Shafer编制的30对双极形容词评定量表<sup>[20]</sup>等,在国内由王登峰等所编制的QZPAS(Chinses Personality Adjective Rating Scale)则采用6点记分<sup>[21]</sup>。

研究者在建构人格测验时都面临着一个同样的问题:为了使测验工具具有较优的测量学性能,测验条目到底应该选用什么样的记分方式?为了解决这一问题测量学家们进行了大量研究,最终还是没有达成一致的观点。出现这种现象可能存在以下原因<sup>[22, 23]</sup>:①在界定测验工具心理测量学属性时所用的标准或准则不同;②考察测验工具的信效度时所运用的心理测量学模型不同。Adelson和McCoach研究发现条目记分的争论主要体现在两方面<sup>[24]</sup>:一是在设置条目记分时是采用迫选还是非迫选;二是条目反应选项的最佳个数。0、1记分就是迫选的代表,在李克特多级记分中迫选与非迫选的区别就在于后者存在中间选项(如“不确定”)。测量学家对这两种形式持有不同的观点和态度,赞同采用非迫选认为这种形式能够更好的识别被试的反应结果,使得测验分数更具有可靠性且被试更喜欢这样的记分方式<sup>[25, 26]</sup>。持相反观点的则认为设置中间选项可能会降低被试对反应选项的辨别,且在这种情形下被试更多时候会选择中间选项。如果采用迫选式,被试会更认真地思考选项和问题进行作答,从而使得评定结果更加准确<sup>[24, 27-29]</sup>。另外,当研究者在开发某测验工具时决定采用迫选式的记分方式,此时究竟采用4点还是6点记分,这就涉及到条目反应选项个数的问题。自李克特记分方式诞生以来,许多研究者一直试图通过考察项目反应个数与内部一致性的关系来确定项目反应选项的最佳个数。然而得出的研究结果却不一致,有时甚至是相互矛盾的。如有研究显示反应选项的个数对内部一致性 $\alpha$ 系数不产生影响<sup>[30, 31]</sup>,与之相反的是有研究显示两者之间存在影响与被影响的关系。在实际应用中,不同研究者推荐的项目反应选项的个数差异很大,少则3点记分,多则达18点记分。如Ko推荐6点或7点记分<sup>[32]</sup>;Preston和Colman研究的记分类型则采用左右两极“刻度”方式,最后推荐7~10点记分<sup>[33]</sup>。

综上所述,目前学术界对人格测验中条目呈现方式和记分方式的效应尚存在争议,但有一点却毋庸置疑,即在建构人格测验时,测验条目选择合适的呈现方式和记分方式可以使得在编制测验时达到事

半功倍的效果,且可得到较优的测量学性能。本研究试图对人格测验中条目的呈现方式和记分方式进行实证研究,以期对实际工作有一定的参考价值。

## 1 对象与方法

### 1.1 对象

研究对象来自深圳和贵阳3所高校在校大学生,采取方便取样共发放问卷800份,400人完成6点记分的单一双极词,另外400人完成7点记分的单一双极词。前者收回有效问卷370份,回收率为92.50%,女生224人,男生134人(未填12人)。后者收回有效问卷364份,回收率为91.00%,女生207人,男生148人(未填9人)。

### 1.2 研究设计

采取2×2研究设计,其中记分方式采取被试间设计即两组被试完成条目的两种记分。呈现方式则采取被试内设计即同一组被试同时完成条目的两种呈现方式。同时采取ABBA设计以平衡顺序效应。以6点记分单一双极词为例,选取一半被试先作答双极词,一周后再作答单极词部分;另外一半被试先作答单极词,一周后再作答双极词部分。

### 1.3 研究工具

1.3.1 自编6点记分的《大五人格双极形容词评定量表》包括5个维度(外倾性、宜人性、严谨性、神经质、开放性),每个维度12个条目,共60题。

1.3.2 自编6点记分的《大五人格单极形容词评定量表》包括5个维度(外倾性、宜人性、严谨性、神经质、开放性),每个维度24个条目,共120题。

1.3.3 自编7点记分的《大五人格双极形容词评定量表》包括5个维度(外倾性、宜人性、严谨性、神经质、开放性),每个维度12个条目,共60题。

1.3.4 自编7点记分的《大五人格单极形容词评定量表》包括5个维度(外倾性、宜人性、严谨性、神经质、开放性),每个维度24个条目,共120题。

### 1.4 共同方法偏差的控制

本研究主要采用问卷法对同一被试进行测试,这容易产生由共同方法变异所产生的共同方法偏差。根据相关研究<sup>[34]</sup>的结果和建议,本研究主要运用以下几种程序修正的方法:①单一双极词的作答方式不一样;②进行匿名调查;③使用反向题目;④采用ABBA以平衡顺序效应。

### 1.5 研究程序

1.5.1 条目的选取 以“大五”人格模型为理论框架,借助专家讨论以及“大五”人格模型各维度的操

作定义,从条目库中选取60个条目(60对双极词如外向的一内向的)组成评定词表,而单极词是双极词的两倍,有120个条目。在选取人格特质形容词时严格按照人格特质词汇选择的标准<sup>[35]</sup>,并参考相关研究<sup>[36,37]</sup>的做法,最后选择出的人格特质词汇均为三字形容词。

1.5.2 施测与数据处理 由心理学专业人士采用统一的指导语,利用上课时间对被试进行团体施测。数据采用SPSS16.0和SAS9.1统计软件相关命令进行处理。

## 2 结 果

### 2.1 分数转换

本研究通过考察因子结构的稳定性,同时辅以信度系数,来探讨条目呈现方式和记分方式的效应。根据相关研究<sup>[14,38]</sup>发现,大五人格模型的因子结构不适宜进行CFA。故本研究中因子结构主要借助于EFA的结果,信度系数则借助于 $\alpha$ 系数和重测信度。根据相关研究<sup>[12]</sup>的建议,对单极词的结果分数进行线性转换时,可以直接运用某一极词的分数减去其所对应的相反极词的分数即可。以6点记分的单一双极词为例,对双极词进行评分,可得到60对双极词的测验分数,用6-60-Bs(Bipolar scores)表示;由于单极词的格式和数目与双极词不相同,在对单极词进行评分时得到120个单极词的测验分数,用6-120-Ss(Single scores)表示;对6-120-Ss进行线性转换时,用某一极的分数减去其对应相反极的分数,最后得到60对差异分数,记为6-60-Ds(Difference scores)。根据Goldberg的研究<sup>[12]</sup>100-TDA和50-Ds得到的因子结构几乎相同,两者对应各维度因子分数的相关均达0.987以上,故将50-Bs和100-TDA的比较直接转变为50-Bs与50-Ds之间的比较。本研究为了便于进行因子结构和信度系数的比较,对单极词也进行类似的处理,得到两组60-Ds,分别是6-60-Ds和7-60-Ds。最后将6-60-Bs与6-120-Ss,7-60-Bs与7-120-Ss的比较转变为6-60-Bs与6-60-Ds,7-60-Bs与7-60-Ds之间的比较。

### 2.2 条目呈现方式

2.2.1 6点记分单一双极词的比较 首先对6-120-Ss和6-60-Ds进行比较,结果发现两组数据得到的相对应各维度因子分数的相关系数分别为外倾性0.970,开放性0.948,严谨性0.971,宜人性0.881,神经质0.808,均达到显著性水平( $P<0.01$ ),6点记分单一双极词的比较只需进行6-60-Bs和6-60-Ds的比

较即可。对6-60-Bs进行探索性因素分析发现,49个双极词(81.67%)在目标因子上的因素负荷最大;相对应的6-60-Ds也进行EFA,结果见表1、表2。

表1结果显示,6-60-Bs中除5个条目的共同度小于0.30以外,其余大部分在0.40~0.60之间;6-60-Ds中有3个条目的共同度小于0.30,其余大部分在0.40~0.60之间。

在确定因子个数时除依据传统方法(如特征值大于1和碎石图检验)外,还借助于平行分析(Parallel analysis)<sup>[39,40]</sup>的结果来确定因子个数。碎石图检验结果表明,在第5个因子之后碎石图的陡度呈平缓,同时平行分析的结果显示,真实数据特征值的碎石图与随机数据平均特征值的曲线相交,且两者交点之上的因子个数也仅有5个。故根据以上标准综合考虑抽取5个公共因子比较合适。对初始因子矩阵进行方差正交旋转以获得因子结构,各条目在对应维度上的因素负荷见表2。表2结果显示,6-60-Bs中所有条目在其目标因子上的因素负荷均超过0.40,且不存在跨负荷现象。而6-60-Ds中有6个条目存在跨负荷现象,分别是V11、V48、V57、V37、V7和V8。

表1 四组数据各变量的共同度

变量	共同度				变量	共同度			
	6-60-Bs	6-60-Ds	7-60-Bs	7-60-Ds		6-60-Bs	6-60-Ds	7-60-Bs	7-60-Ds
V1	0.603	0.610	0.700	0.685	V15	0.390	0.484	0.484	0.532
V2	0.674	0.546	0.651	0.663	V45	0.533	0.537	0.625	0.615
V11	0.300	<b>0.165</b>	0.423	0.383	V46	0.736	0.705	0.746	0.784
V13	0.367	0.307	0.364	0.364	V35	0.625	0.475	0.665	0.642
V31	0.628	0.543	0.470	0.662	V36	0.511	0.514	0.552	0.618
V32	0.623	0.635	0.716	0.708	V26	0.329	0.371	0.515	0.467
V21	0.620	0.630	0.603	0.681	V55	0.643	0.633	0.753	0.737
V22	0.527	0.601	0.662	0.696	V56	0.350	0.392	0.445	0.483
V42	0.626	0.522	0.680	0.585	V7	0.670	0.759	0.653	0.568
V51	0.552	0.525	0.602	0.579	V8	<b>0.298</b>	0.467	0.508	0.527
V52	0.592	0.760	0.705	0.801	V17	0.413	0.389	0.444	0.489
V41	0.364	0.594	0.518	0.649	V37	0.559	0.642	0.684	0.664
V9	<b>0.297</b>	0.348	0.403	0.466	V28	0.512	0.514	0.471	0.514
V10	0.458	0.475	0.609	0.673	V47	0.455	0.612	0.512	0.599
V19	0.384	0.452	0.597	0.461	V48	0.718	0.756	0.602	0.728
V20	<b>0.296</b>	0.359	0.326	0.399	V57	0.457	0.303	0.606	0.432
V30	0.493	0.432	0.634	0.635	V58	0.446	0.618	0.559	0.641
V39	0.729	0.516	0.711	0.696	V14	<b>0.277</b>	0.395	0.445	0.508
V40	0.414	0.417	0.491	0.535	V23	0.332	0.369	0.538	0.556
V49	0.426	0.478	0.352	0.394	V33	0.362	<b>0.170</b>	<b>0.268</b>	<b>0.228</b>
V50	0.593	0.664	0.688	0.672	V4	0.374	0.304	0.454	0.449
V59	0.523	0.550	0.499	0.509	V3	0.465	0.403	0.633	0.603
V60	0.501	0.622	0.613	0.586	V53	<b>0.275</b>	<b>0.227</b>	0.513	0.457
V5	0.596	0.546	0.618	0.714	V54	0.393	0.311	0.484	0.518
V6	0.571	0.567	0.734	0.670					

注:共同度小于0.30者被加粗。

2.2.2 7点记分单一双极词的比较 采取相同方法对7-120-Ss和7-60-Ds的结果进行比较,发现两组数据得到的相对应各维度因子分数的相关在0.852~0.951之间,均达到显著性水平( $P<0.01$ )。随后对7-60-Bs和7-60-Ds进行EFA,结果见表1、表2。表1



显示,7-60-Bs和7-60-Ds中仅V39的共同度小于0.30,其余条目的共同度均在0.40~0.60之间。根据碎石图检验和平行分析的结果显示,抽取5个公共因子比较合适。对初始矩阵进行方差正交旋转,各

条目在对应维度上的因素负荷见表2。表2显示,7-60-Bs中所有条目(除V13外)在其目标因子上的因素负荷均超过0.40,且不存在跨负荷现象。而7-60-Ds中神经质维度的全部条目均存在跨负荷现象。

表2 四组数据的因子结构

变量	F1 外倾性				F2 开放性				F3 严谨性				F4 神经质				F5 宜人性			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
V42	0.748	0.664	0.744	0.633			0.318	0.330												
V21	0.714	0.754	0.674	0.703			0.364	0.357												
V2	0.728	0.60	0.760	0.754													0.318			
V1	0.713	0.663	0.772	0.685													-0.402	0.306		
V31	0.678	0.587	0.543	0.626													-0.452	0.350		
V32	0.674	0.724	0.726	0.764														0.361		
V52	0.615	0.797	0.648	0.808				0.323					-0.397		-0.409					
V51	0.564	0.578	0.564	0.489			0.337	0.337			0.349	0.434								
V41	0.499	0.524	0.489	0.637														0.484		
V22	0.492	0.679	0.551	0.696							0.390	-0.385		-0.473						
V11	0.402	0.225	0.510	0.145																
V39					0.839	0.667	0.784	0.774												
V50					0.743	0.799	0.804	0.770												
V30					0.695	0.650	0.782	0.747												
V60					0.666	0.736	0.762	0.726												
V10					0.661	0.649	0.771	0.781												
V59					0.634	0.631	0.671	0.674												
V19					0.604	0.649	0.732	0.602												
V40					0.582	0.547	0.647	0.662												
V49					0.547	0.565	0.514	0.451				0.307								
V9				0.314	0.493	0.416	0.554	0.477				0.333								
V20		0.301			0.477	0.504	0.515	0.580												
V46									0.771	0.811	0.788	0.786					0.316			
V55				0.334					0.749	0.770	0.805	0.723								
V35									0.738	0.649	0.719	0.730								
V6			0.335						0.725	0.713	0.770	0.773								
V5									0.721	0.694	0.716	0.751								
V45									0.682	0.648	0.707	0.749								
V36									0.581	0.429	0.693	0.725		-0.385				0.346		
V15									0.560	0.573	0.675	0.725						0.360		
V56									0.525	0.512	0.651	0.667								
V26									0.489	0.507	0.658	0.642								
V28		-0.478		<b>-0.538</b>								-0.365	0.688	0.492	0.531					
V48	-0.546	<b>-0.790</b>	-0.329	<b>-0.786</b>									0.648	0.341	0.657					
V57		<b>-0.389</b>		<b>-0.474</b>									0.646		0.673					
V47		-0.374		-0.432				-0.414	-0.343	-0.380	<b>-0.479</b>	0.632	0.563	0.512						
V37	-0.399	<b>-0.696</b>	-0.361	<b>-0.739</b>								0.630	0.387	0.726						
V7	-0.536	<b>-0.836</b>		<b>-0.730</b>								0.615		0.740						
V58		-0.328		-0.435		-0.391	-0.409		-0.338	-0.334	<b>-0.524</b>	0.613	0.573	0.530						
V17		-0.348		-0.314						-0.323	<b>-0.506</b>	0.594	0.462	0.577						
V8		<b>-0.463</b>		<b>-0.561</b>							-0.340	0.512		0.683				-0.360		
V3																	0.607	0.570	0.731	0.690
V4																	0.590	0.547	0.666	0.619
V54																	0.556	0.468	0.592	0.663
V33																	0.547	0.364	0.427	0.349
V13		<b>0.384</b>															0.542	0.484		0.411
V23																	0.523	0.529	0.642	0.688
V53																	0.512	0.472	0.680	0.665
V14															-0.333		0.483	0.590	0.548	0.548

注:1表示6-60-Bs,2表示6-60-Ds,3表示7-60-Bs,4表示7-60-Ds;因素负荷小于0.3者省略;跨因子负荷者被加粗。

2.3 条目记分方式

2.3.1 因子结构 在条目呈现方式的比较中,无论是6点还是7点记分,双极词的因子结构均好于单极词的结果。确定条目呈现方式之后,接着对6点和7

点记分双极词的因子结构与信度系数进行比较。结果见表1。表1显示,6-60-Bs和7-60-Bs中绝大多数条目的共同度在0.40~0.60之间。根据碎石图检验和平行分析的结果,抽取5个因子比较合适,进行

方差正交旋转,两组数据中各条目在对应维度上的因素负荷见表2。表2显示,7-60-Bs中所有条目(除V13外)在其目标因子上的因素负荷均超过0.40,且V13存在跨因子负荷现象。而6-60-Bs中所有条目在其目标因子上的因素负荷均超过0.40,且不存在跨负荷现象。就因子结构的清晰性而言,6-60-Bs的结果也较优于7-60-Bs。

2.3.2 信度系数 表3结果显示,6-60-Bs和7-60-Bs中各维度的 $\alpha$ 系数均在0.85以上(除宜人性外),平均分别为0.84和0.86。同时间隔10周的重测信度显示6-60-Bs和7-60-Bs的重测系数均大于0.70。

表3 6-60-Bs和7-60-Bs的信度系数

信度系数	测验工具	外倾性	宜人性	严谨性	神经质	开放性	Mean
$\alpha$ 系数	6-60-Bs	0.87	0.75	0.89	0.85	0.86	0.84
	7-60-Bs	0.90	0.78	0.91	0.87	0.85	0.86
10周重测	6-60-Bs(73)	0.79	0.74	0.82	0.71	0.78	0.77
	7-60-Bs(69)	0.82	0.76	0.79	0.72	0.77	0.77

注:括号内为重测人数。

### 3 讨 论

#### 3.1 条目呈现方式

通过分别比较6点和7点记分单一双极词的因子结构,结果表明:无论6点还是7点记分,双极词的因子结构均优于单极词形式。究其原因可能确实与单一双极词的呈现形式有关系,从条目的构成上讲,单极词实际上只呈现了一个词汇,而双极词呈现的是一组词,且彼此互为参照,被试在作答时更容易的理解问题<sup>[11,12]</sup>。尽管有学者认为运用双极词时可能会使得每对双极词的两端都被强制看成是相反的<sup>[11]</sup>,但我们认为如果在进行条目的筛选时考虑到被试的教育水平和实际情况,再综合条目的熟悉度和意义度等,这种情况可以避免。Goldberg<sup>[12]</sup>认为使用双极词更能清楚地描述量表所要测量的内容和维度,但他在实证研究时却发现单极词的因子结构要较优于双极词,单极词中98%的变量在目标因子上的因素负荷最高,而在双极词测验中只有82%的变量在目标因子的因素负荷最高。本研究所得结果与Goldberg<sup>[12]</sup>的研究结果不一致,可能的原因是:其一由于中英文语言形式在表达同一件事情方式的差异;其二因为中英文表达反义词构词方法的差异,英文中许多反义词是通过单词前面加上否定的前缀如dis-,un-等构成反义词,在Goldberg<sup>[12]</sup>研究中确实也大量采用了这样的方式来组成双极词条目。而本研究选择双极词时特别参考了中文《近义词反义词词典》等,所有双极词均是由两个完全不同的形容词所

组成,因此这些条目算是真正意义上双极词。

#### 3.2 条目记分方式

本研究结果表明6-60-Bs的因子结构较优于7-60-Bs,其原因可能是:①与6点记分相比,7点记分中增加了中间选项,被试在作答时可能会陷入犹豫不决的情形,从而影响其测量的结果<sup>[28,29]</sup>;②受中国传统文化中庸之道的的影响,部分被试在作答时倾向于选择中间的答案。王孟成等<sup>[19]</sup>在编制CBF-PI时也发现类似的情况。故采取迫选方式对被试者来说测量效果会更好,因为此时被试不得不对问题进行认真的回答以做出明确的选择<sup>[27]</sup>。

从内部一致性系数来看,6-60-Bs各维度的 $\alpha$ 系数略低于7-60-Bs。Adelson和McCoach的研究<sup>[24]</sup>也曾发现5点记分得到的信度系数要高于4点记分。可能的原因是:①与6点记分方式相比,7点记分增加了一个评定等级,从而增大了每个条目和量表得分的全距,这会对信度系数产生一定的影响;②与6点记分相比,在7点记分中各选项之间的距离变短,使得条目间的同质性增强,也可能导致 $\alpha$ 系数的增高;③7点记分中选择中间项的被试较多,使得分数的散点分布更接近纺锤形,从而提高了测验信度。

重测信度是用来衡量测验工具的稳定性指标,对于人格测验显得非常重要。从理论上讲个体的人格特征具有跨时间、跨情景的相对稳定性,在心理测量中将前后两次实测数据之间的相关大小作为衡量测验跨时间的稳定性。故本研究在考察信度系数时也报告了间隔10周的重测信度。结果发现:6-60-Bs和7-60-Bs的重测信度均达到较好水平,各因子的重测信度均大于0.70,这也从旁佐证了大五人格模型的跨时间、跨情境稳定性的假设。

最后,尽管两种记分方式的因子结构与信度系数结果存在轻微的不一致,我们仍然倾向于选择6点记分方式。原因如下:首先,在评价一个测验的测量学属性时主要还是根据测验的效度而非信度,因为信度高是效度高的必要非充分条件。Cronbach<sup>[25]</sup>指出除非测验的效度也得到提高,否则仅仅提高测验的信度是毫无意义可言的。Lei Chang<sup>[22]</sup>认为在评价量表的记分方式时测验效度是更好的评价标准。其次,中国人的行为方式深受“中庸”文化背景的影响,当遇到难以决策的情况时往往倾向于选择中间选项,但被试者过多地选择中间项实际上提示他们没有对测验情境表明态度,从而降低了测验的效度。第三,从测量效果角度考虑,设置中间选项的记分方式有可能会降低被试者对量表反应选项的辨别

动机,被试者更多的时候可能会选择中间选项。相反去掉中间选项即采用迫选式的话,被试可能会更认真地思考问题和选项进行作答,从而使得评定结果更准确<sup>[24,27-29]</sup>。

### 参 考 文 献

- John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In John OP, Robins RW, Pervin LA(Ed.). *Handbook of Personality: Theory and Research*(3th). New York: The Guilford Press, 2008. 114-158
- 罗杰,戴晓阳. “大五”人格测验在我国使用情况的元分析. *中国临床心理学杂志*, 2011, 19(6): 740-742
- Weems GH, Onwuegbuzie AJ, Lustig D. Profiles of respondents who respond inconsistently to positively-and negatively worded items on rating scales. *Evaluation and Research in Education*, 2003, 17(1): 45-60
- 郭庆科, 韩丹, 王昭, 等. 人格测验中题目正反向陈述的效应. *心理学报*, 2006, 38(4): 626-632
- Mook J, Kleijn WC, Ploeg HM. Positively and negatively worded items in a self-report measure of dispositional optimism. *Psychological Reports*, 1992, 71(1): 275-278
- Samuelstuen MS. Psychometric properties and item-keying direct ion effect s for the Learning and study strategies inventory-high school version with Norwegian students. *Educational and Psychological Measurement*, 2003, 63(3): 430-446
- Bae SW, Brekke JS. The measurement of self-esteem among Korean Americans: A cross-ethnic study. *Cultural Diversity and Ethnic Minority Psychology*, 2003, 9(1): 16-33
- 王孟成, 蔡炳光, 吴艳, 等. 项目表述方法对中文 Rosenberg 自尊量表因子结构的影响. *心理学探新*, 2010, 30(3): 63-68
- Stansbury JP, Ried LD, Velozo CA. Unidimensionality and bandwidth in the center for epidemiologic studies depression (CES-D) scale. *Journal of Personality Assessment*, 2006, 86(1): 10-22
- Goldberg LR, Kilkowski JM. The prediction of semantic consistency in self-descriptions: Characteristics of persons and of terms that affect the consistency of responses to synonym and antonym pairs. *Journal of Personality and Social Psychology*, 1985, 48(1): 82-98
- Duijsens IJ, Diekstra RW. The 23BB5: a new bipolar big five questionnaire. *Personality and Individual Differences*, 1995, 19(5): 753-755
- Goldberg LR. The development of markers for the big-five factor structure. *Psychological Assessment*, 1992, 4(1): 26-42
- Caprara GV, Barbaranelli C, Borgogni L, et al. The big five questionnaire: a new questionnaire to assess the five factor model. *Personality and Individual Differences*, 1993, 15(3): 281-288
- 戴晓阳, 姚树桥, 蔡太生, 等. NEO 个性问卷修订本在中国的应用研究. *中国心理卫生杂志*, 2004, 18(3): 170-174
- 戴晓阳, 吴依泉. NEO-PI-R 在 16-20 岁人群中的应用研究. *中国临床心理学杂志*, 2005, 13(1): 14-18
- Lee K, Ashton MC. Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 2004, 39(2): 329-358
- 王登峰, 崔红. 中国人人格量表(QZPS)的编制过程与初步结果. *心理学报*, 2003, 35(1): 127-136
- 王登峰, 崔红, 胡军生, 等. 中国青少年人格量表(QZPS-Q)的编制. *心理发展与教育*, 2006, (22)3: 110-115
- 王孟成, 戴晓阳, 姚树桥. 中国大五人格问卷的初步编制 I: 理论框架与信度分析. *中国临床心理学杂志*, 2010, 18(5): 545-548
- Shafer AB. Brief bipolar markers for the five factor model of personality. *Psychological Reports*, 1999, 84(3): 1173-1179
- 崔红, 王登峰. 中国人人格形容词评定量表(QZPAS)的信度、效度与常模. *心理科学*, 2004, 27(1): 185-188
- Chang L. A Psychometric evaluation of 4-point and 6-point likert-type scales in relation to reliability and validity. *Applied Psychological Measurement*, 1994, 18(3): 205-215
- Alberto MO, Kramp U, Carlos GF, et al. The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*, 2009, 41(2): 295-308
- Adelson JL, McCoach DB. Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point likert-type scale. *Educational and Psychological Measurement*, 2010, 70(5): 796-807
- Cronbach LJ. Further evidence on response sets and test design. *Educational and Psychological Measurement*, 1950, 10(1): 3-31
- Ory JC, Wise SL. Attitude change measured by scales with 4 and 5 response options. Paper Presented at the Meeting of the National Council on Measurement in Education, Ghicago, IL, 1981
- Garland R. The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 1991, 2(1): 66-70
- Busch M. Using likert scales in L2 research: A researcher comments. *TESOL Quarterly*, 1993, 27(4): 733-726
- Reid J. The dirty laundry of ESL survey research. *TESOL Quarterly*, 1990, 24(2): 323-338
- Aiken LR. Number of Response Categories and Statistics on a Teacher Rating Scale. *Educational and Psychological Measurement*, 1983, 43(2): 397-401